

Contour Hierarchies, Tied Parameters, Sound Models and Music

Lonce Wyse

Institute for Infocomm Research, Singapore
ponce@zwhome.org, www.zwhome.org/~ponce

Abstract

Our goal is to construct sound generating model structures that capture relationships among sound components that are perceived by human listeners as musical when the model parameter space is traversed. Designing constructs that are in certain ways homologous to perceptual organization, results in sound model structures that are not only possible to exploit for “expressive” performance, but that can play a direct role in the compositional listening strategies for an audience. A case study of a model of a Canyon Wren song is used to illustrate the modeling principles.

1 Sound model design

A generative sound model has three components: a) a range of sounds, b) a set of exposed parameters, c) behavior that specifies how the space of sounds is traversed under parameter manipulation. The process of sound model design frequently starts with a specification of the range of sounds and some constraints on the behavior. Sometimes a composer may have actual samples lying within the target range of sounds, but needs a model in order to generate a broader class of sounds that includes those in the specification as a special case, and/or because of a need for effective “handles” for moving around the space of sounds.

The hierarchical structures, contours, parameter mappings and tyings, and relationships among sound components embodied in an algorithm provide the definition and character of a virtual sound source that the listener can use in their musical listening strategies. For example, knowing the range of sounds and behaviors of a model sets the conditions for expectations and “surprise” that have been so much the topic of literature on musical listening (Meyer, 1956). Once a listener is familiar with the sound models that are being used in a composition, they can be used in the cognitive organization of unfolding sonic material. This is particularly important in electroacoustic music where a shared body of knowledge about harmony and melody are not available to help in organizing the listening experience.

In physical sound modeling (Smith 2002, Cook 2002a), the structural constraints are taken from the properties of materials such as string, tubes and plates. Physical models generally expose parameters that are intuitive, easy to learn how to control, and whose effects on the sound are easy to perceive given the shared knowledge that we all have about the physical world.

It is not only that the constraints are physical that make these models work. Indeed, it is commonly noted that with physical models we can do things that would be impossible in the real world, such as putting a vibrato on material thickness. Thus it is not the physical plausibility *per se* of these models that make them so intuitive and valuable in a musical context. It is the very fact that there are constraints and structure in the model that gives us the impression of a well defined sound source, even if a real physical source cannot be identified as the sound generator.

With acoustic modeling, we don’t have, or don’t use, pre-made structure, but use only the sound as a guide to model structure. There are theoretically an infinite number of model architectures capable of generating a given set of sounds (though in practice, it may be difficult to find even one). The challenge is to find structure – relationships between component features that give clear character and definition to a perceived sound source, even without having to hear the whole range of sounds it can make. If models are “strong” in this sense, then it should be easy to tell, for example, whether a given sound comes from a certain model.

There are several ways that we can build structure “behind” the surface representation of a sound. One way is to analyze a sound into a set of dynamic acoustic (e.g. spectral) features, and then attempt to reduce the redundancy in our representation using some variant of Principle or Independent Components Analysis. One of the objectives of this approach is to come up with a small number of parameters that represent a sound example and a class of sounds in a “neighborhood” of the target sound. Another goal is to discover a low dimensional set of parameters that are “perceptually significant”.

We cannot expect such automatic redundancy discovery to always find the structure that we so easily hear in a sound. The following example

illustrates the construction of hierarchical contours and parameter tying in creating a model of the Canyon Wren song.

2 The Canyon Wren

The Canyon Wren song exhibits a canonical structure with a series of events that change systematically over its course. In keeping with the terminology in the literature on bird song analysis (e.g. Leonardo and Konishi, 1999), we refer to the whole signal shown in Figure 1 as a “motif”, and the individual chirps as “syllables” characterized by being separated by silence or an abrupt change in acoustic features. A layer above the motif is called a ‘bout’, and generally exhibits characteristic structure as well, but we only consider the motif here.

From the time series view of the audio signal (Figure 1), we can clearly see an amplitude contour of the motif, a slowing down of the syllable event rate over the course of the motif, and a change in the per-syllable amplitude over the course of the motif.

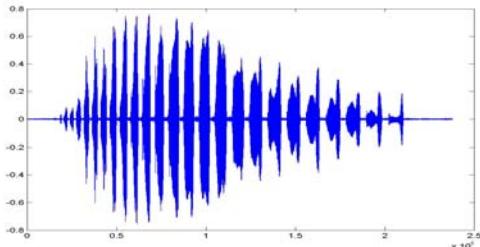


Figure 1. The time domain signal of the Canyon Wren song exhibits structure on multiple time scales.

The frequency shape and structure of individual syllable events are more clearly visible in the sonogram (Figure 2). Here we see that there is a systematic change in the individual event frequency contours from something of a “seahorse” shape with a cutoff tail at the beginning of the motif, to the full seahorse shape in the middle, and to one with a cutoff tail and flattened back at the end. Furthermore, we see that the number of harmonics visible in this representation grows from none at the beginning, peaks at 3 roughly one third of the way through the gesture, and trails off to none again by the end.

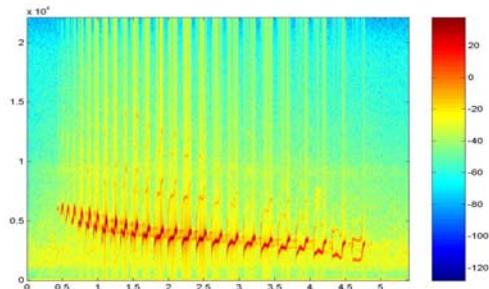


Figure 2. The sonogram makes more multiscale structure visible.

The coarse features we have discussed for the Canyon Wren song motif can be described by a

handful of “musical gestures”, or “contours”. One is a broad unimodal contour with a peak about one third of the way through the gesture (Figure 3) that corresponds to shape of the amplitude and the harmonic content of the motif.

A second contour is approximately linear from the beginning to end. The most obvious feature controlled by the linear contour in this model is the center frequency of the individual syllables over the duration, which can be seen in Figure 2.

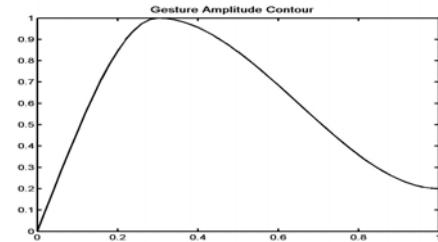


Figure 3. A motif-level contour.

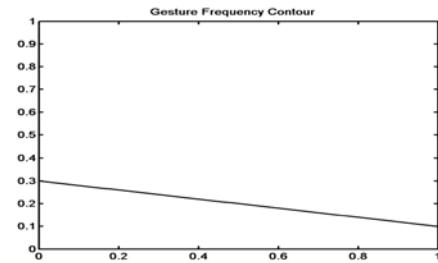


Figure 4. A second motif-level contour, independent of the first.

Two other contours are used at an entirely different time scale, that of the individual syllables. One has the shape of a “seahorse” (Figure 5), and the other similar, but with a flattened back. (Figure 6). These contours are also visible in the sonogram of Figure 2 as the individual syllables at the beginning and at the end of the motif.

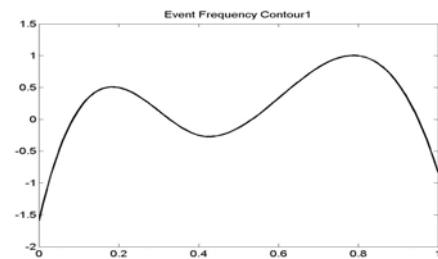


Figure 5. An event-level contour, visible in the syllable frequency shapes at the beginning of the motif.

Contour shapes are represented in convenient units (0 centered, or zero minima) and map the ranges to appropriate units for specific acoustic attributes. The ranges used in mapping contours to attributes or other internal parameters can themselves be exposed as model parameters.

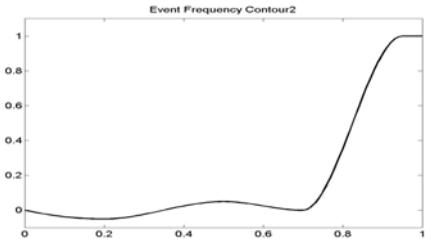


Figure 6. Another event-level contour, visible in the syllable frequency shapes at the end of the motif.

3 Parameter tying

Tying internal parameters has the effect of reducing the control space dimensions, and creating an invariant characteristic across a possibly wide range of sounds a model might produce. Figure 7 shows how the various contours control the sound synthesis for the Canyon Wren model.

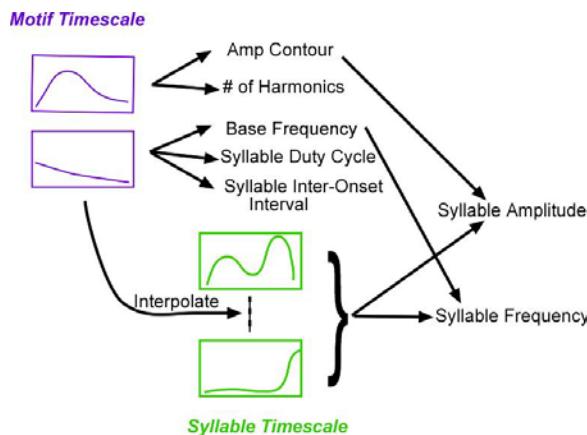


Figure 7. Hierarchy of contours and tied parameter relationships defining the character of the Canyon Wren song.

The contour hierarchy results in an interrelated, time-evolving, many-to-one control over the acoustic attributes of amplitude and frequency for the syllables. Not visible in the figure are the remaining independent parameters that the model exposes for external control. External controls include one that determines the time rate of progression through the top-level contours, and several others that determine the range or baseline values of an attribute (e.g. syllable rate) to which the contours map. They can all be manipulated in real-time as the motif plays. This model can be heard and explored at the URL, <http://www.zwhome.org/~asw/ASWExp/CanyonWren/ASWApplet.html>.

Manipulating the exposed parameters does not change the qualitative structural relationships defined by parameter tying. The preserved structure tends to give a perceptual quality of invariance that psychologists use to describe the unity of percept across surface variation that is so important for object perception. This kind of object perception can be a tremendously powerful musical concept that unifies

quite disparate sounds. This is similar to the role played by the systematic relationship between pitch and a wide variety of timbres in musical instrument perception. Similarly, structural “landmarks” in melodies afford the unifying perception of transformed material as themes and variations in traditional tonal music. (Carterette, Kohl and Pitt, 1986).

4 Synthetic Canyon Wren Results

The time domain and sonogram of the synthetic Canyon Wren using “best fit” parameters for the model can be seen in Figures 8 & 9, respectively.

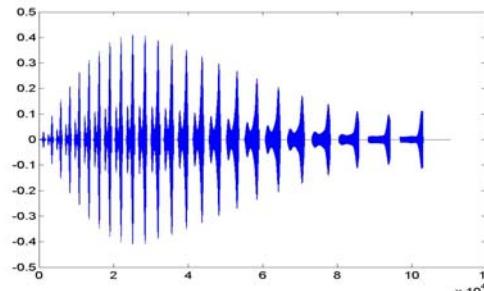


Figure 8. Time domain signal of the resynthesized motif.

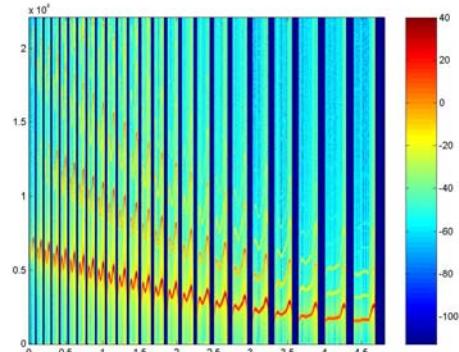


Figure 9. Sonogram of the resynthesized motif.

The “synthetic” nature of the signal is visible. Nonetheless, the subtlety of the evolution of the sound, despite its simplicity, results in a lifelikeness that comes through clearly in listening to the sounds. Furthermore, the model has “integrity” in that transformations can be made via the exposed parameters without destroying the perceptibility of the structural invariants that define the sound source. Themes and variations can be generated that bear the same kind of “family resemblance” relationship to each other that melodic themes and variations do in traditional music. Models can thus provide the connection among sequences of sounds in electroacoustic music that melodic and harmonic transformations provide in traditional music. A “model based listening” approach unifies theories of

musical listening across both traditional and less sonically constrained music.

5 Discussion

The Canyon Wren model takes on its defining characteristics from the shapes and hierarchical control structure of the contours. The contours are tightly integrated in the processes of signal and event generation that are internal to the model. The “synthesis algorithm” (to make a somewhat artificial distinction) is simply a sum of three sine waves, and labeling the control structures a “parameter mapping” glosses over their essential contribution to the model. In fact, due to the event patterns and higher order parameter control structure in the model, a wide variety of different possible synthesis algorithms can be used while preserving the identity of the model.

The process of building the Canyon Wren model involved taking a single example of a sound (the bird song) and producing a model capable of generating a large class of perceptually related sounds that includes the example sound at a particular parameter setting. Better computational support for this process would be valuable for composers and sound designers, and is the focus of ongoing work.

One reason the process is difficult to automate is that the characteristics that give a model its uniqueness are far removed from an objective signal level description of a sound. For example, recognizing each chirp syllable in the Canyon Wren song as a parameterizable variant of an underlying shape is an interpretive act, not a matter of extracting information that is “in” the signal. Nonetheless, a well-designed tool that involves the sound designer in segmenting the signal, customizing signal analysis routines, building appropriate mathematical models of feature contours and incorporating them into synthesis structures could significantly reduce the amount of time it takes to create a synthesis model such as the one described in this paper.

The distinction between “instrument” and “composition” is not always clear or helpful in contemporary music. This is particularly true in music free of any *a priori* constraints on the domain of sound from which it draws its source material. There is commonly no simple relationship between the physical gestures and the resulting sound as there is in traditional musical instrumental music. It can happen, for example, that many sonic events might be the result of a single physical gesture. In these contexts, it becomes useful to conceptualize mapping structures, event patterns, and control contours together as part of a single sound model. Musical sound modeling is thus akin to interactive sound effects development for games and virtual worlds where the domain includes any and all sound and there are a diverse set of interaction demands on sound synthesis (Wyse and Kellock 1999, Cook 2002b).

6 Conclusion

The structures and mappings described here are part of the general process of sound modeling and sound composition. The parameter mapping and tying and the hierarchical control structure in models create high order invariance structure that is persistent across sound transformations generated by the model under parameter variation. Invariance structure gives identity and character to an object, allowing a listener to build up an “internal” model of a sound source (that may be physical or otherwise). We used abstract contours and hierarchies in the construction of a Canyon Wren model which are analogous to contours and structures that are formed based on perceptual laws of organization such as proximity, similarity, closure and simplicity. Listening activities such as hierarchisation, abstraction and simplification of representation form an important part of musical listening and meaning formation, although these mechanisms have usually been discussed in relation to melody (Deutsch, 1982).

It is interesting to consider that with the aid of computers, composers can build explicit generative structures that used to be only implicit in the surface level of their musical scores. The relationship between model structures and disparate sounds across a piece make them potentially very useful artifacts for the analysis of contemporary music.

References

- Carterette, E., Kohl, D., & Pitt, M., 1986. Similarities among transformed melodies: The abstraction of invariants. *Music Perception*, 3, 393-410.
- Cook, P., 2002a. *Real Sound Synthesis for Interactive Applications*. AK Peters, Natick MA.
- Cook, P., 2002b. Sound Production and Modeling. *IEEE Computer Graphics and Applications*. July/August 2002, 23-27.
- Deutsch, D., 1982. The Processing of Pitch Combinations. In D.Deutsch (Ed.) *The Psychology of Music*. New York Academic Press: NY.
- Koffka, K. 1935. *Principles of Gestalt Psychology*. NY, NY: Harcourt, Brace and Company.
- Leonardo, A. and Konishi, M., 1999. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* **399**, 466 – 470.
- Meyer, L. B., 1956. *Emotion and Meaning in Music*. Chicago: Chicago University Press.
- Smith, Julius O., 2002. *Digital Waveguide Modeling of Musical Instruments*. Center for Computer Research in Music and Acoustics (CCRMA), Stanford University. www-ccrma.Stanford.edu/~jos/waveguide/
- Wyse, L. and P. Kellock, 1999. Embedding Interactive Sounds in Multimedia Applications, *Multimedia Systems Journal*, 7:1, 48-54.